Intra-disciplinary differences in database coverage and the consequences for bibliometric research

Tove Faber Frandsen, Jeppe Nicolaisen

Intra-disciplinary differences in database coverage and the consequences for bibliometric research*

Tove Faber Frandsen† and Jeppe Nicolaisen

Royal School of Library and Information Science

Birketinget 6, DK-2300 Copenhagen S., DENMARK

tff@db.dk; jni@db.dk

Abstract: Bibliographic databases (including databases based on open access) are routinely used for bibliometric research. The value of a specific database depends to a large extent on the coverage of the discipline(s) under study. A number of studies have determined the coverage of databases in specific disciplines focusing on inter-disciplinary differences. However, little is known about the potential existence of intra-disciplinary differences in database coverage. Focusing on intra-disciplinary differences, the paper documents large database coverage differences within two disciplines (economics and psychology). The point extends to include both the uneven coverage of specialties and research traditions. The implications for bibliometric research are discussed, and precautions which need to be taken are outlined.

Introduction

The introduction of large bibliographic databases marks a significant development in the history of bibliometrics. Many branches of bibliometric research have grown out of, or been made possible by the use of these databases. However, the use of bibliographic databases for bibliometric research is not without its problems. According to Hood and Wilson (2003: 593), these problems may be seen as falling within three categories:

1. Errors or lack of consistency in the data (at the micro level)
2. Other types of problems and difficulties (at the macro level)
3. Problems with the tools that are made available by the database provider or host

One of the problems dealt with in the second category, is that of database coverage. This problem has both quantitative and qualitative aspects (Jacsó, 1997). The quantitative aspects concern among other things the size of the database(s), indexed document types, the number of English-language and foreign-language source documents, geographic coverage, and the time span and currency of the database(s). The qualitative aspects are partly about the inclusion of core journals and prestigious non-journal sources.

Using bibliographic databases for bibliometric research implies using these databases and their coverage as censuses of publications comparable to demographers using population censuses for demographic studies (White & McCain, 1989). Potential bias in the censuses will reflect itself in the results of studies based on these. Thus, the coverage of bibliographic databases has consequences for bibliometric research and this is valid regardless of specific choice of database (subject-specific databases, citation databases or databases based on open access resources).

It is crucial to recognize possible coverage problems before conducting bibliometric research. Using the citation indexes for research evaluation may produce quite biased and invalid results. Moed (2005) discusses the coverage problem and its implications for the use of the citation databases produced by the Institute of Scientific Information (ISI) in research evaluation. He presents a so-called *tentative* classification of disciplines according to their overall ISI coverage into three classes, with *excellent, good,* and *moderate coverage*, respectively. Disciplines in the *"excellent"* category include molecular biology and biochemistry, biological sciences primarily related to humans, clinical medicine, physics and astronomy, and chemistry. The *"good"* category contains the disciplines of applied physics and chemistry, biological sciences primarily related to animals and plants, psychology & psychiatry. As well as other social sciences primarily related to medicine and health, geosciences, mathematics, engineering, and economics. The *"Moderate"* category contains other social sciences, and humanities and arts. Moed's classification system gives one the impression that whole disciplines are either excellent, good, or moderately covered in the citation databases, thus making research evaluation based on ISI data feasible to a varying degree. Although it is just a tentative classification system, it is nevertheless still too crude,

as a division of disciplines into smaller units may reveal uneven coverage and thus casting the idea about *the* coverage of whole disciplines in to doubt. Unfortunately, there seems to be a lack of studies of the bibliographic coverage of specialties and research traditions within disciplines and its consequences for bibliometric research. This paper aims to close this gap. The focus is centered on intra-disciplinary differences. More specifically, the paper aims to investigate the bibliographic coverage of specialties and research traditions within disciplines and the consequences for bibliometric research. As such, the paper to some extent continues in the footsteps of previous studies. Yet, its narrow focus on intra-disciplinary differences distinguishes it from related studies focusing on, for instance, inter-disciplinary coverage, geographical coverage and coverage of publication types.

The paper is organized as follows: The next section provides an overview of related research. The following sections present the methods of data gathering and processing followed by results, discussions, and conclusions.

Related research

Disciplinary differences in publishing and citing behavior have been studied in various ways, but only seldom at the intra-disciplinary level. A search for studies focusing on database coverage issues at the intra-disciplinary level yielded no results. Yet, subject-specific database comparisons have been undertaken in a number of fields.

Inter-disciplinary and intra-disciplinary differences

Differences in publishing behavior can be analyzed on several levels. The analysis can be performed on a macro level as a comparison between e.g. the humanities, social sciences, natural sciences, and medicine as done by Kyvik (2003). The analysis can also be made at a more detailed level as done by Hyland (2000) examining the relationships between the cultures of eight disciplines and their unique discourses. Kling and McKim (2000) examine the heterogeneity of communications illustrated by an analysis of three disciplines. Knievel and Kelly (2005) compare eight humanities fields. Kyvik (1988) focuses on six social sciences as he compares them with other fields and analyzes the differences among the social sciences. Lindholm-Romantschuk and Warner (1996) study the role of monographs in scholarly communication in philosophy, sociology and economics. Metcalfe (1995) illustrates the differences

between disciplines by showing a difference in mean publication lag between two disciplines of 6.2 and 16.3 months. Nederhof et al. (1989) and Nederhof and Zwaan, (1991) analyze the composition of document types, their coverage by the citation indexes and the consequences for bibliometric research in six disciplines. An even more fine-grained analysis can be performed as shown by Bordons and Zulueta (1997) stressing that even within the same ISI heading differences are found between journals representing different research communities. Pharmacological teams and pharmacy teams are identified and their results show that the journals they submit their articles to for publication are very different. Hamilton (1990, 1991) shows that the un-citedness rate varies from 36.7 to 88.0% among fields and from 9.2 to 99.8 among sub-disciplines indicating that inter-disciplinary differences are not necessarily larger than intra-disciplinary differences. Laband (2002) compares conditions of co-authorships in economics and agricultural economics uncovering great differences in the author conditions.

Database coverage and bibliometric research

A number of researchers have investigated the coverage of a bibliographic database and the consequences for bibliometric research. The bibliometric consequences related to the problem of geographical coverage are well illustrated by Webster's (1998) analysis of a Polish sociological citation index (PSCI) and the Social Science Citation Index (SSCI). Her findings strongly imply that bibliometric indicators based on SSCI paint one picture of Polish sociology and the PSCI another. Another study by Narvaez-Berthelemot and Russell (2001) contains an analysis of 4,326 social science journals from the UNESCO DARE-database. It reveals that 64% of production of journals in the world takes place in high income countries. Furthermore, that SSCI primarily consists of journals from the rich countries (97%). Apart from that there is a smaller group of journals from middle income countries and finally there is a very small group of journals from low income countries (0.7%). All countries apart from the US have fewer journals included in SSCI than the UNESCO DARE-database. Bordons, Fernandez and Gomez (2002) report on some of the problems for peripheral countries in relation to calculations of journal impact factors, and stress that it should be borne in mind that large parts of the scientific output in these countries are not included in the citation indexes. Studies on of database coverage are also made using open access data sources. Hajjem, Harnad and Gingras (2005) analyze coverage of 1,307,038 articles from 10 disciplines from 1992-2003 by open access based resources and find an overall percentage of OA

articles ranging from 5 to 16% depending on discipline, year and country. Swan et al. (2005) have studied the coverage of different document types by open access resources and find some document types to be better covered by open access resources. The bibliometric consequences related to the problem of document type coverage are evident in the study by Cronin, Snyder and Atkins (1997). The three authors constructed a database comprising 30,000 references from 90 books randomly chosen among those reviewed in top sociological journals. They compare lists of the 26 authors most cited in the books and in the top 24 sociology journals, and their findings demonstrate that there are two distinct populations of highly cited authors in sociology: One consisting of authors cited in the journal literature, another of authors cited in the monographic literature. Given the citation databases' limited coverage of monographic citing material, the latter population may regularly go unrecognized. Finally, the number of databases needed to cover the literature on a specific topic has been found to vary considerably. Hood and Wilson (2001) report that for a typical search topic, the single most inclusive database covers 23 to 37% of the relevant literature. To cover 95% of the literature in their study of 14 topics would require the use of 11 to 35 databases. Hood and Wilson (2001) study topics from various fields. Other researchers have examined database coverage of particular subject areas. Walters and Wilder (2003) provide a comprehensive overview of this research.

Methods

The focus of this study is on the coverage of bibliographic databases and the consequences for bibliometric research with a narrow focus on the disciplines of economics and psychology. Both are classified as well-covered in Moed's classification system (Moed, 2005: 138), however, a more detailed analysis is needed. This may be accomplished by dividing the disciplines into research traditions or specialties. All disciplines embrace a number of (often competing) research traditions that to some extent are distributed among the specialties that shape the whole discipline. What characterizes a specialty is, according to Meadows (1998), the phenomenon or phenomena, which members of the specialty study. Laudan (1977) invokes the idea of a large-scale unit in science that he calls a *research tradition*. A research tradition is held together by common ontological assumptions about the nature of the world and methodological principles about how to revise theories and develop new ones.

Research traditions are consequently not the same as specialties. A research tradition is "a set of ontological and methodological do's and don'ts" (Laudan, 1977: 80) whereas a specialty is a specific part, fraction or division of a larger discipline.

The specialties within the discipline of economics were determined using EconLit. EconLit is the American Economic Association's electronic bibliography of economics literature. EconLit contains abstracts, indexing, and links to full-text articles in economics journals. It abstracts books and indexes articles in books, working papers series, and dissertations. EconLit indexes the economics literature using EconLit Subject Descriptors, which is comparable to headings in the JEL Classification System (www.econlit.org). The JEL Classification System is a classification developed for the economics literature by the Journal of Economic Literature and widely used in the discipline. Barrett, Olia and Von Bailey (2000) also use the JEL Classification System to show that economics is a discipline characterized by great specialization. Other databases could have been used as the benchmark database, but EconLit was chosen because of the subject descriptors.

The year 1991 was chosen as the starting point for the analyses because the required information was not available for the previous years. A 15 year publication period (1991-2005) was employed. On the basis of the JEL classification system the following four specialties were selected: Health economics, mathematical and quantitative methods, economic history and schools of economic thought and methodology. The varying publication patterns were analyzed using the JEL Classification System in EconLit. All publications indexed with the classification code of the specialty were ranked according to publication type year by year. An overview of the document composition is available in appendix 1 to 4. The same publications were also ranked according to journal name, resulting in 15 annual lists for each specialty of journals publishing one or more articles indexed in EconLit with the classification code of the specialty. Subsequently, the 60 lists of journals resulting in a total of 34,496 journal articles were scrutinized and checked for indexing in the citation databases (Social Sciences Citation Index, Science Citation Index, Arts & Humanities Citation Index) for each of the examined years. The investigation was made on journal level and not article level, implying that each article was not looked up in the indexes, but the journal was. This implies that if only a selection of the articles in a journal is indexed in the citation indexes, it is possible that the specific article is not indexed but as the citation indexes normally

include all research articles of a journal (Moed, 2005: 113) the possible bias of this procedure is assessed to be low.

To give a preliminary answer to the problem concerning the consequences of coverage, a study of the relative sizes of the four specialties was conducted. The relative size of specialties was analyzed using different sources for performing research evaluation. As this analysis includes open access sources an analysis of the 15 years is not possible because the content of open access based databases is not static and thus the most recent year in the analysis (2005) was selected. The point of reference was EconLit, the citation databases and a delineation in the citation databases to the top 20 journals within the economics subject category. The top 20 journals were measured by journal impact factor (JIF) as available through the 2005 JCR social sciences edition in the subject category economics (isiknowledge.com/jcr). An overview of the 20 journals is available in appendix 5. The rationale for the analysis performed on the 20 journals with the highest JIF is a number of previous studies that have used a similar delineation as the sampling method (e.g. Hodgson & Rothman, 1999; Kocher & Sutter, 2001; Frost et al., 2003) or as means to characterize the quality of publications (Klaić & Klaić, 2004). Furthermore, a tool for citation analysis based on open access resources was included. The publications were located using Google Scholar - an alternative to the existing citation databases (Noruzi, 2005; Bakkalbasi et al., 2006; Kousha & Thelwall, 2006; Neuhaus & Daniel, 2007). Only journal articles were included in order to make the analysis comparable to the one in the citation databases. Unlike the citation indexes, open access based resources do not allow systematic analyses of neither the indexing policy nor the consequences of it. The indexing policies of services based on open access resources are difficult to analyze. This is partly caused by a lack of available information of the indexing policy (e.g. Google Scholar) and partly because the service providing access to the data is not in control of the indexing policy (e.g. archives based on authors self-archiving their work). Consequently, the influence on bibliometric studies of the indexing policy is even more difficult to investigate as we are limited by the options available through the services and as Neuhaus and Daniel (2007) state:

> "Google Scholar currently processes its sources in an unsystematic, unpredictable and fragmentary manner. For lack of adequate options for browsing, searching and saving

> results in structured output formats it is difficult to make even elementary bibliometric analyses efficiently."

Consequently, all 4,230 journal articles had to be looked up individually in Google Scholar. Noruzi (2005) outlines the search techniques available through Google Scholar. In this case the queries submitted were based on words from the title and author's last name. If the query did not yield any results, the number of words were initially increased and afterwards decreased. It had to be a total match to be registered as available in full text via Google Scholar. This implies that a journal article had to be available as pre-print or post-print. This could lead to a decrease in the shares of publications retrieved as OA. But there is a risk of vast differences between e.g. a working paper and the subsequent journal article so the distinction had to be made. Furthermore, it was checked if the publication was available in full text or if only the bibliographic information was available. Some links are to toll access journals and others to open access databases that may not include access to full text. RePEc (Research Papers in Economics) is an example of a decentralized database of working papers, journal articles and software components freely available. However, as stated on the website (http://repec.org), RePEc does not contain full-text journal articles as RePEc services provide links to many full text articles. Yet, a personal or institutional subscription to follow those links is often required.

In order to be able to study differences at the level of research traditions, three research traditions were chosen from the psychological specialty psychotherapy and psychotherapeutic counselling:

- cognitive therapy,
- behavior therapy
- psychoanalytic therapy.

These three are different research traditions because they hold different ontological assumptions about psychological phenomena as well as different ideas about how to study them (Robins, Gosling & Craik, 1999; Nicolaisen, 2004: chapter 5). To assess the variations in coverage of various databases caused by intra-disciplinary differences, the specialties within the selected discipline of psychology must be determined. For that purpose we used the database PsycINFO that indexes the literature in psychology

and related behavioral and social sciences, including psychiatry, sociology, anthropology, education, pharmacology, and linguistics. Records from 1967 and beyond are indexed using the controlled vocabulary from the Thesaurus of Psychological Index Terms.[1] The publications of three research traditions were determined using the Thesaurus of Psychological Index Terms. An overview of the composition of document types and the total publications is available in appendix 6 to 8. Note that due to the indexing policy of PsycINFO the shares of working papers are not available in these appendices.

The same 15 year publication period (1991-2005) was employed, and the varying publication patterns were analyzed using the Thesaurus of Psychological Index Terms in PsycINFO. This resulted in 45 lists of journal names comprising 16,193 publications. All publications indexed with the classification code of the research tradition were ranked according to publication type year by year. The same publications were also ranked according to journal name. Again, the list of journals was scrutinized and checked for indexing in the citation databases. The study of the relative sizes of the three research traditions was performed using the same method as the study of economics specialties. The only difference was the list of top 20 journals measured by JIF which was determined by merging the 10 subject categories related to psychology. An overview of the 20 journals is available in appendix 9. In Google Scholar all 1,366 articles were looked up individually.

Results

In the following we show the development in the size of a selection of document types. Each figure depicts the moving averages[2] of a specific document type of the total research output within a specialty from 1991 to 2005 in intervals of three years (although the first and last year are only in intervals of two years).

Figure 1 shows the significance of journal articles within the four selected specialties. It should be noted that this document type includes all types of journal articles (e.g. reviews, research articles and notes).

[1] The three index terms used are *cognitive therapy, behavior therapy & behavior modification*, and *psychoanalytic therapy*.

[2] A moving average is simply the average of a series of numbers over a period of time which is constantly updated by dropping the oldest value and then adding the newest value and recalculating the average. Using moving averages smooth a data series and make it easier to spot trends.

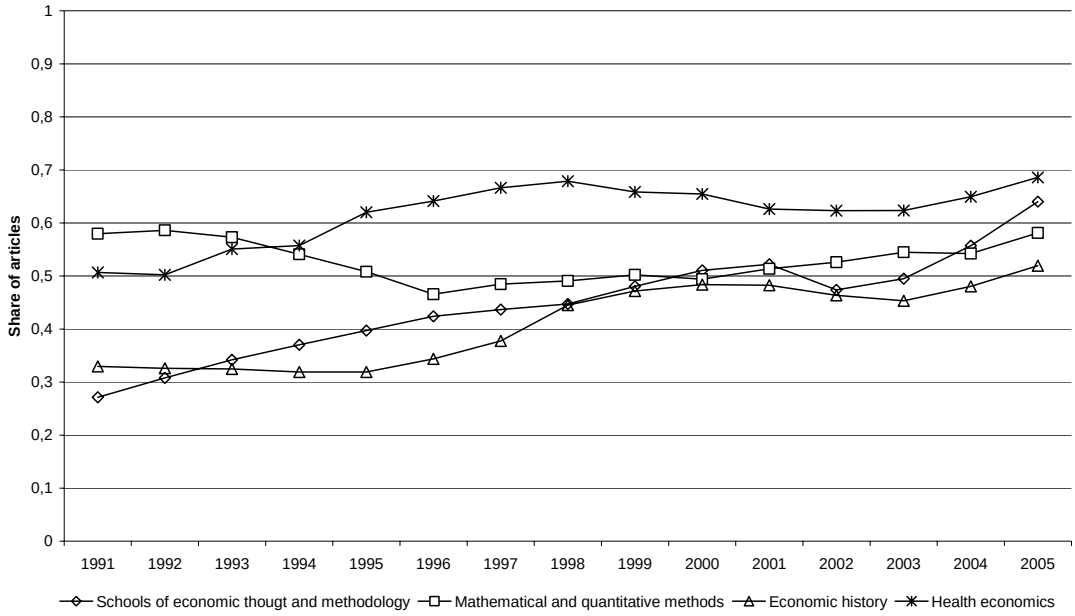

Figure 1. Share of journal articles of total output in four economics specialties in EconLit 1991-2005

As is quite evident, the journal article is of growing importance in all the specialties, and it increases from shares of 25 to 56% in 1991 to shares of 57 to 71% in 2005. The remaining document types are primarily books and working papers. The results showing the importance of these document types are available in appendix 1-4. The relative size of the book seems to be relatively stable in some disciplines on a level of about 2 to 7% of the research output. However, within one specialty it appears as if books are losing their importance. In the specialty of economic history the book is rapidly decreasing in relative size over the years, although, the book is still at a much higher level in this specialty than in the other three. The working paper is a document type with an increasing significance within all four specialties. However, the importance of the working papers is varying considerably among the specialties as mathematical and quantitative methods hold a share of 30% in 2005 whereas the other specialties have shares varying from 4 to 12% in 2005.

Based on this analysis we can conclude that specialties within the discipline of economics have quite varying publication patterns, and we will now examine the implications of these findings for the coverage in the citation databases. Figure 2 illustrates the coverage of journal articles in the citation indexes.

Figure 2 depicts the shares of journal articles covered by the citation indexes. It should be noted that the influence of publication patterns is excluded as only journal articles are included. The coverage varies from 40 to more than 90% of the journal articles.

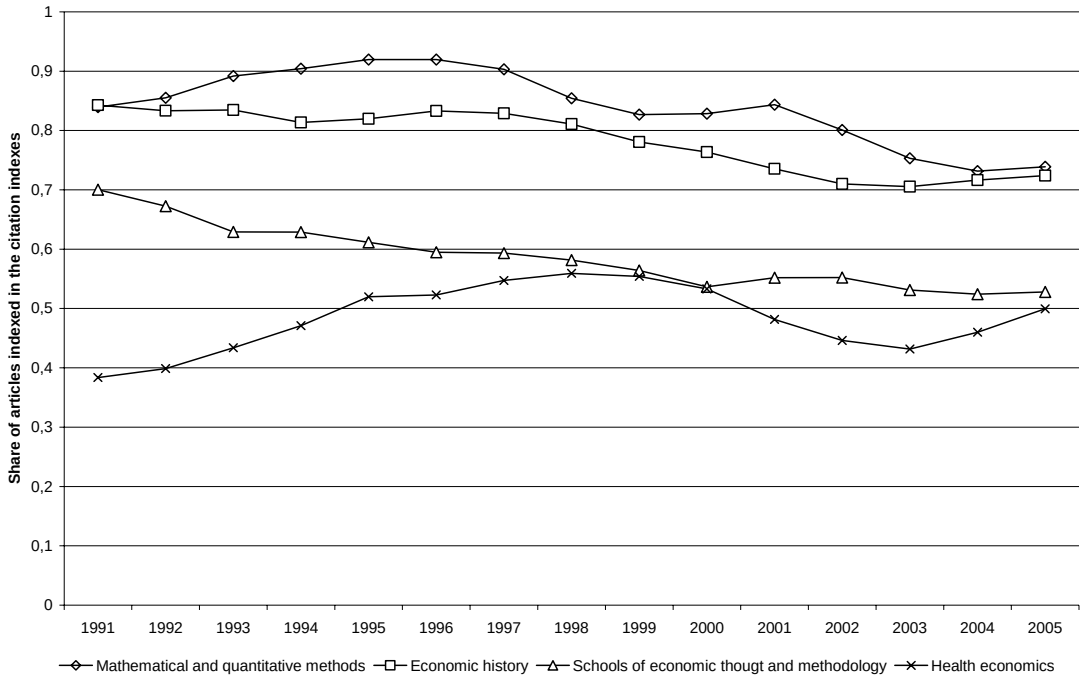

Figure 2. Share of journal articles indexed in the citation databases from 1991 to 2005 of four economics specialties.

Obviously, in a research evaluation these varying degrees of coverage can hypothetically imply that some specialties appear less productive than others. However, the central issue is to what extent the varying degrees of coverage influence the results of citation analyses and research evaluation. To give a preliminary answer, we have conducted a study of the relative sizes of the four specialties in 2005. The results are presented in table 1.

	EconLit	ISI citation databases	Top 20 journals	Google Scholar
Mathematical and quantitative	36	38	21	50

methods				
Economic history	17	17	13	19
Schools of economic thought and methodology	19	15	07	13
Health economics	28	30	60	18

Table 1. Relative sizes of four economics specialties in per cent: journal articles published in 2005.

Table 1 shows the relative sizes of the four specialties vary considerably when using different pools of documents. It should be noted that there is no "true" relative size among these four pools of documents as they are all determined by their indexing policy. In EconLit mathematical and quantitative methods make up 36% of the total amount of journal articles produced by these four specialties. Economic history is the smallest amounting to 17%. Schools of economic thought & methodology and health economics are represented by respectively 19 and 28%. In an evaluation performed using EconLit of productivity measured by the number of journal articles this would be their relative sizes. The same analysis done by using the citation databases would depict a somewhat different picture. Economic history would hold the same relative size whereas health economics and mathematical and quantitative methods would have slightly bigger shares. However, this increase in size is associated with a decrease for schools of economic thought and methodology which would appear to be a less productive area than e.g. economic history although in EconLit it was the other way around. Turning to the 20 journals with the highest JIF, the relative sizes are considerably different from the two previous pools of documents. The share of health economics doubles (and becomes the largest specialty) and the rest of the specialties lose shares (although they do not lose shares equally). Finally, if the analysis had been performed using Google Scholar, health economics turns into one of the three smallest specialties whereas mathematical and quantitative methods becomes the dominating specialty by far. Summing up the table, it is evident that these four pools of documents are not duplicating the same picture of productivity in these four economics specialties.

Taking the analysis one step further, we analyze three research traditions within one specialty as we look at three research traditions of psychology: Cognitive therapy, behavior therapy & behavior modification and psychoanalytic therapy.

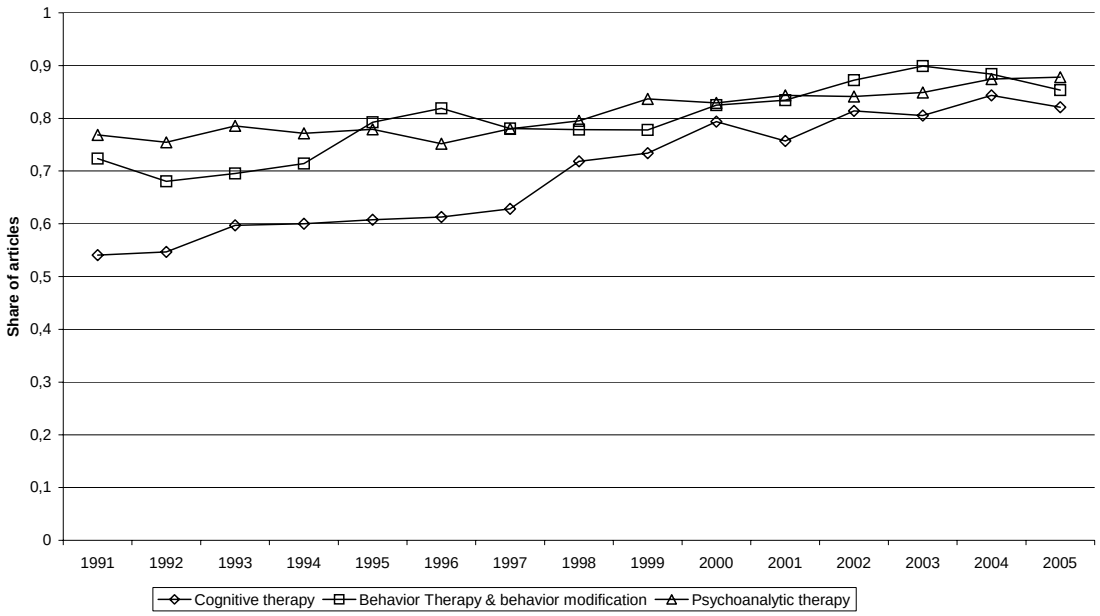

Figure 3. Share of journal articles in three Psychological research traditions in PsycINFO 1991-2005.

Figure 3 shows the significance of journal articles within the three selected research traditions. Again, it should be noted that this document type includes all types of journal articles (e.g. reviews, research articles and notes). Compared to the four economics specialties the journal article is much more important and is also of growing importance in all the research traditions as it increases from shares of 51% to 74% in 1991 to shares of 81% to 91% in 2005.

The three research traditions have relatively similar publication patterns, during the last 5 or 6 years of the period and the publication patterns cannot be used to explain differences in visibility. Their visibility in a research evaluation is to a large extent dependent upon the indexing policy of the tools used for the research evaluation. As can be seen in figure 4, the coverage of journal articles in ISI varies considerably across research traditions and these three research traditions are thus not indexed equally well each year.

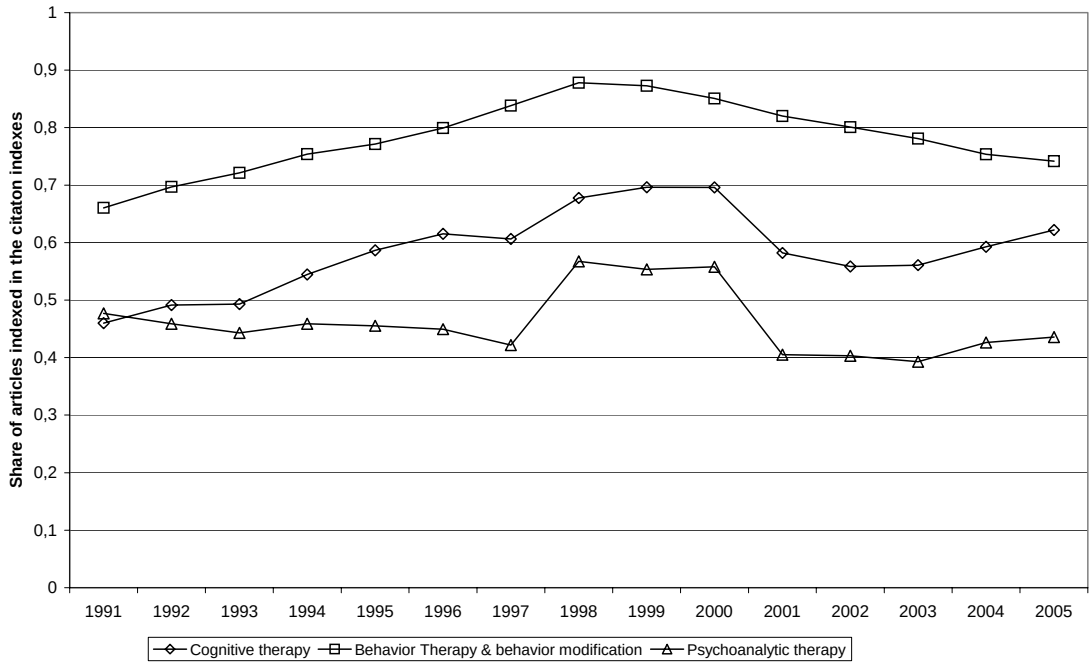

Figure 4. Share of journal articles indexed in the citation databases from 1991 to 2005 of three Psychological research traditions.

Throughout the entire period, behavior therapy & behavior modification is considerably better covered by the citation indexes compared to the other two research traditions in general, and psychoanalytical therapy in particular. The poor coverage of the latter is somewhat surprising as this research tradition has its own subject category in the citation indexes (*Psychoanalysis*).

Turning to the implications of the uneven coverage of research traditions, table 2 provides an overview of the relative sizes of the three research traditions.

	PsycINFO	ISI citation databases	Top 20 journals	Google Scholar
Cognitive therapy	31	36	69	41
Behavior therapy & behaviour modification	14	20	28	22

Psychoanalytical therapy	54	44	03	37

Table 2. Relative sizes of three Psychological research traditions in per cent: journal articles published in 2005.

Like the specialties within economics, the relative sizes of the three research traditions also vary considerably using different pools of documents. Again it must be stressed that there is no "true" relative size among these four pools of documents. In PsycINFO psychoanalytical therapy provides a little over 50% of the journal articles of these three research traditions. Cognitive therapy produces 31% of the journal articles and behavior therapy & behaviour modification the remaining 14%. The same analysis, using the citation databases as pool of documents, produces a different picture. Psychoanalytical therapy loses 10 percentage points whereas the other two research traditions share the "profit" almost equally. However, the ranking of research traditions by size does not change. The same cannot be said about the relative sizes using the 20 highest ranking JIF journals. Psychoanalytical therapy almost vanishes and cognitive therapy becomes the clearly identifiable leading research tradition. Finally, using Google Scholar involves lost shares for psychoanalytical therapy reducing it to the second largest research tradition. Using Google Scholar does, however, depict a picture of two strong research traditions and a third somewhat smaller research tradition. The problems of uneven coverage of specialties within economics are therefore also evident when it comes to research traditions in psychology. The four pools of documents are not replicating the same picture of productivity in the three Psychological research traditions.

Discussion

The results of our empirical study show that a number of specialties in the discipline of economics and research traditions in the discipline of psychology are not represented equally well in the databases. As the findings of the present study only relate directly to the disciplines of economics and psychology, we cannot assume they aply to other disciplines. However, though restricted to these disciplines the results do have broader implications.

Using a different method, Moed (2005) evaluates the coverage of the citation indexes in order to lay the groundwork for a proper understanding of the performance measures computed on the basis of ISI data. His study includes an analysis of the coverage of economics in which he finds the percentage of references to documents published in journals, relative to total references in 2002 to be 56 (Moed, 2005: 129). Furthermore, he finds the ISI coverage of journals within economics to be 83% leading to an overall ISI coverage of 47. However, these figures can be further differentiated as we find the importance of the journal as a publishing medium to range from 52 to 63% of the total output. Furthermore, we find the ISI coverage of journal articles to range from 58 to 83%. This leads to an overall ISI coverage of 30, 36, 46 and 48%. This variation of values indicates that some specialties within economics are what Moed refers to as well covered although others are only moderately covered.

In psychology and psychiatry Moed (2005: 130) finds the percentage of references to documents published in journals, relative to total references in 2002 to be ranging from 69 to 81 (psychology and psychiatry is divided into sub-disciplines). Furthermore, he finds the ISI coverage of journals to be 86 and 91% leading to an overall coverage in psychology and psychiatry of 60 to 73%. However, these figures can be further differentiated when looking at research traditions within psychology as we find the importance of the journal as publishing medium to range from 74 to 86% of the total output. Furthermore, we find the ISI coverage of journal articles to range from 39 to 85%. This leads to overall coverage of 33, 41 and 73%. In the three examined research traditions the variation is even greater than in the four economic specialties. Some psychological research traditions are well covered whereas others are just moderately covered.

According to Moed (2005: 140), the degree of coverage of a field determines the type of research assessment study necessary to perform an adequate analysis. The moderately covered fields should not be analysed relying on ISI data alone, but require supplementary analyses based on data not available in the ISI databases. In some of the moderately covered fields it may not even be possible to perform citation analyses. The research assessment study needs to be adjusted according to the field as studies based solely on ISI data at risk of being biased in moderately covered fields.

It is easy to imagine how bibliometric studies based on an uneven coverage of a discipline's specialties and research traditions, could produce biased or invalid results. Normally a distinction is made between

two kinds of bibliometric studies. The first concerns studies based on publication analysis. The second concerns studies based on citation analysis.

Publication analyses normally seek to measure and compare the scientific output of authors, institutions and countries. This is usually accomplished by counting the number of publications indexed in databases. It is thus of utmost importance that databases used for publication analyses cover all specialties and research traditions of the analyzed disciplines adequately. Otherwise the bias in the coverage will immediately reflect itself in the results of the publication analysis, thus invalidating its conclusions.

Bias will also reflect itself in the results of citation analyses. There are four main applications of citation analysis (Zunde, 1971; Nicolaisen, 2007):

1. Qualitative and quantitative evaluation of scientists, publications and scientific institutions
2. Modeling of the historical development of science and technology
3. Information search and retrieval
4. Knowledge organization based on bibliographic coupling and co-citation analysis

Authors tend to cite authors from the same specialty and/or research tradition (Nicolaisen, 2004). Uneven database coverage of specialties and research traditions will consequently affect the results of all four applications. The volume of citations to the well covered specialties and research traditions will be disproportionate higher than the volume of citations to the ill covered specialties and research traditions.

Another problem with uneven database coverage of a discipline's specialties and research traditions concerns the issue of sampling. The majority of bibliometric studies are based on retrieved data from databases. The databases are normally used for two related purposes: 1. for selecting a sample for further analysis, and 2. for detecting the publication output of the sample units and/or how many times the sample units are cited. Blind reliance on uneven database coverage, when selecting a sample for further analysis, is clearly problematic. Such a sample may at best be regarded a fractionized sample, and any results based on such a sample has limited generalizability (Nicolaisen, 2006).

When conducting bibliometric studies it is crucial to identify possible coverage problems that may lead to biased results. To recognize such problems the analyst must be knowledgeable about the discipline(s) under study. It is vital to be aware of various specialties and research traditions within the discipline(s),

and to examine their coverage in the databases selected for studies. It may be possible to compensate for uneven database coverage, but only if the analyst knows *what* to normalize for.

Conclusion

Intra-disciplinary differences in database coverage affect the results of bibliometric research based on retrieved data from databases. We have documented significant differences in the disciplines of economics and psychology, and revealed quite uneven coverage of economic specialties and psychological research traditions. These observable facts have consequences for all bibliometricians - not only those studying the disciplines of economics and psychology. Intra-disciplinary differences in database coverage could very well be found in other disciplines as well. Consequently, specialties and research traditions of any discipline cannot be assumed to be covered equally well in the databases. It is important to be aware of this and to take appropriate precautions before initiating bibliometric studies using bibliographic databases.

Acknowledgements: The authors wish to thank the anonymous referees for their constructive suggestions for improvements and Peter Joseph Woods for proof-reading the manuscript.

References

Bakkalbasi, N.; Bauer, K.; Glover, J. & Wang, L. (2006). Three options for citation tracking: Google Scholar, Scopus and Web of Science. http://eprints.rclis.org/archive/00006080/02/GS_Scopus_WoS_04182006_preprint.pdf. [Downloaded 05-30-2007].

Barrett, C.B; Olia, A. & Von Bailey, D. (2000). Subdiscipline-specific journal rankings: Whither Applied Economics. *Applied Economics*, 32: 239-252.

Bordons, M.; Fernandez, M.T. & Gomez, I. (2002). Advantages and limitation in the use of impact factor measures for the assessment of research performance in a peripheral country. *Scientometrics*, 53(2): 195-206.

Bordons, M. & Zulueta, M. A. (1997). Comparison of research team activity in two biomedical fields. *Scientometrics*, 40(3), 423-436.

Cronin, B., Snyder, H. & Atkins, H. (1997). Comparative citation rankings of authors in monographic and journal literature: A study of sociology. *Journal of Documentation,* 53(3): 263-273.

Frost, S.; Murphy, R.; Webster, P. & Schmidt, U. (2003). Are top journals biased against eating disorders topics? *American Journal of Psychiatry*, 160(2), 363–365.

Hajjem, C., Harnad, S. & Gingras, Y. (2005). Ten-Year Cross-Disciplinary Comparison of the Growth of Open Access and How it Increases Research Citation Impact. IEEE Data Engineering Bulletin 28(4), 39-47.

Hamilton, D. P. (1990). Publishing by - and for? - the numbers. *Science,* 250, 1331-32

Hamilton, D. P. (1991). Research papers: who's uncited now. *Science*, 251, 25.

Hodgson, G. & Rothman, H. (1999). The editors and authors of economics journals: a case of institutional oligopoly? The Economics Journal, 109(2), 165-186.

Hood, W.W. & Wilson, C.S. (2001). The scatter of documents over databases in different subject domains: How many databases are needed? *Journal of the American Society for Information Science*, 52(14): 1242-1254.

Hood, W.W. & Wilson, C.S. (2003). Informetric studies using databases: Opportunities and challenges. *Scientometrics*, 58(3): 587-608.

Hyland, K. (2000), Disciplinary Discourses: Social Interactions in Academic Writing, Harlow: Longman.

Jacsó, P. (1997). Content evaluation of databases. *Annual Review of Information Science and Technology*, 32: 231-267.

Klaic, Z. B. & Klaic, B. (2004). Croatian scientific publications in top journals according to the Science Citation Index for the 1980-2000 period. *Scientometrics*, 61(2), 221-251.

Kling, R. & McKim, G. (2000). Not Just a Matter of Time: Field Differences in the Shaping of Electronic Media in Supporting Scientific Communication. *Journal of the American Society for Information Science*, 51(14), 1306-1320.

Knievel, J.E., & Kelley, C. (2005). Citation analysis for collection development: A comparative study of eight humanities fields. *The Library Quarterly* 75(2): 142-168.

Kocher, M & Sutter, M. (2001). The Institutional concentration of authors in top journals of Economics during the last two decades. *Economic Journal*, 111(5), 405-421.

Kousha, K. & Thelwall, M. (2007). Google Scholar Citations and Google Web/URL Citations: A Multi-Discipline Exploratory Analysis. *Journal of the American Society for Information Science and Technology*, 58(7), 1055-1065.

Kyvik, S. (1988). Internationality of the social sciences: the Norwegian case. *International Social Science Journal*, 115, 163-172.

Kyvik, S. (2003). Changing trends in publishing behaviour among university faculty, 1980-2000. *Scientometrics*, 58(1), 35-48.

Laband, D. N. (2002). Contribution, attribution and the allocation of intellectual property rights: economics versus agricultural economics. *Labour economics,* 9, 125-131.

Laudan, L. (1977). *Progress and its Problems: Toward a Theory of Scientific Growth*. Berkeley, CA: University of California Press.

Lindholm-Romantschuk, Y. & Warner, J. (1996), The role of monographs in scholarly communication: an empirical study of philosophy, sociology and economics. *Journal of Documentation*, 52(4), 389–404.

Metcalfe, N. B. (1995). Journal impact factors. *Nature*, 77, 260-261

Moed, H.F. (2005). *Citation Analysis in Research Evaluation*. Dordrecht, NL: Springer.

Narvaez-Berthelemot, N. & Russell, J.M. (2001). World Distribution of Social Science journals: A view from the periphery. *Scientometrics,* 51(1): 223-239.

Nederhof, A. J., Zwaan, R.A., De Bruin, R.E., & Dekker, P. J. (1989). Assessing the usefulness of bibliometric indicators for the humanities and the social sciences: A comparative study. *Scientometrics,* 15, 423-435.

Nederhof, A. J. & Zwaan, R. A. (1991) Quality judgments of journals as indicators of research performance in the humanities and the social and behavioral sciences. *Journal of the American Society for Information Science and Technology*, 42(5, 332-340.

Neuhaus, C. & Daniel, H-D. (2007). Data sources for performing citation analysis: An overview. Accepted for publication in *Journal of Documentation*. http://e-collection.ethbib.ethz.ch/show?type=bericht&nr=490&part=text. [Downloaded 06-01-2007].

Nicolaisen, J. (2004*). Social Behavior and Scientific Practice – Missing Pieces of the Citation Puzzle* [PhD-thesis]. Copenhagen, DK: Royal School of Library and Information Science.

Nicolaisen, J. (2006). Traditional author co-citation analysis: A discussion of the sampling problem. *Proceedings of the 1st International Conference on Multidisciplinary Information Sciences and Technologies, InSciT2006:* 635-639.

Nicolaisen, J. (2007). Citation analysis. *Annual Review of Information Science and Technology*, 41: 609-641.

Noruzi, A. (2005). Google Scholar: The new generation of citation indexes. *LIBRI*, 55(4): 170-180. http://eprints.rclis.org/archive/00005595/01/Google_Scholar%2C_The_New_Generation_of_Citation_Indexes.pdf

Robins, R.W., Gosling, S.D. & Craik, K.H. (1999). An empirical analysis of trends in Psychology. *American Psychologist*, 54(2): 117-128.

Swan, A., Needham, P., Probets, S., Muir, A., Oppenheim, C., O'Brien, A., Hardy, R. and Rowland, F. (2005). Delivery, Management and Access Model for E-prints and Open Access Journals within Further and Higher Education. Technical Report, JISC, HEFCE. http://eprints.ecs.soton.ac.uk/11001/01/E-prints_delivery_model.pdf.

Walter, W. H. & Wilder, W. (2003). Bibliographic index coverage of a multidisciplinary field. *Journal of the American Society for Information Science and Technology*, 54(14): 1305–1312.

Webster, B.M. (1998). Polish sociology citation index as an example of usage of national citation indexes in scientometric analysis of social sciences. *Journal of Information Science,* 24(1): 19-32.

White, H.D. & McCain, K.W. (1989). Bibliometrics. *Annual Review of Information Science and Technology*, 24: 119-186.

Zunde, P. (1971). Structural models of complex information sources. *Information Storage and Retrieval*, 7: 1-18.

Appendix 1. Publications of the specialty: Schools of economic thought and methodology.

Year	Total publications	Journal articles		Books		Working papers		Other publication types	
1991	1687	427	(25)	124	(7)	6	(0)	1130	(67)
1992	1644	476	(29)	116	(7)	2	(0)	1050	(64)
1993	1243	473	(38)	75	(6)	7	(1)	688	(55)
1994	1462	521	(36)	110	(8)	2	(0)	829	(57)
1995	1263	472	(37)	87	(7)	9	(1)	695	(55)
1996	1245	574	(46)	80	(6)	12	(1)	579	(47)
1997	1303	570	(44)	75	(6)	12	(1)	646	(50)
1998	1525	628	(41)	101	(7)	14	(1)	782	(51)
1999	1336	659	(49)	80	(6)	13	(1)	584	(44)
2000	1290	692	(54)	69	(5)	26	(2)	503	(39)
2001	1447	726	(50)	87	(6)	18	(1)	616	(43)
2002	1112	588	(53)	78	(7)	20	(2)	426	(38)
2003	1228	480	(39)	68	(6)	24	(2)	656	(53)
2004	1238	700	(57)	67	(5)	27	(2)	444	(36)
2005	1131	809	(72)	69	(6)	50	(4)	203	(18)

Percentages are shown in brackets.

Appendix 2. Publications of the specialty: Mathematical and quantitative methods

Year	Total publications	Journal articles		Books		Working papers		Other publication types	
1991	1609	902	(56)	71	(4)	67	(4)	569	(35)
1992	1685	1009	(60)	60	(4)	115	(7)	501	(30)
1993	1643	984	(60)	58	(4)	158	(10)	443	(27)
1994	1714	894	(52)	70	(4)	70	(4)	680	(40)
1995	1765	886	(50)	58	(3)	217	(12)	604	(34)
1996	2428	1217	(50)	65	(3)	396	(16)	750	(31)
1997	3080	1214	(39)	102	(3)	408	(13)	1356	(44)
1998	2478	1386	(56)	73	(3)	480	(19)	539	(22)
1999	2755	1431	(52)	74	(3)	488	(18)	762	(28)
2000	3114	1332	(43)	88	(3)	882	(28)	812	(26)
2001	2603	1392	(53)	81	(3)	462	(18)	668	(26)
2002	2016	1167	(58)	69	(3)	408	(20)	372	(18)
2003	1742	809	(46)	62	(4)	528	(30)	343	(20)
2004	2460	1456	(59)	52	(2)	597	(24)	355	(14)
2005	2664	1521	(57)	77	(3)	824	(31)	242	(9)

Percentages are shown in brackets.

Appendix 3. Publications of the specialty: Economic history

Year	Total publications	Journal articles		Books		Working papers		Other publication types	
1991	1331	413	(31)	161	(12)	14	(1)	743	(56)
1992	1103	385	(35)	148	(13)	19	(2)	551	(50)
1993	1232	393	(32)	173	(14)	21	(2)	645	(52)
1994	1304	399	(31)	204	(16)	29	(2)	672	(52)
1995	1319	438	(33)	198	(15)	31	(2)	652	(89)
1996	1479	472	(32)	182	(12)	48	(3)	777	(53)
1997	1410	536	(38)	194	(14)	49	(3)	631	(45)
1998	1399	607	(43)	189	(14)	49	(4)	554	(40)
1999	1186	619	(52)	140	(12)	39	(3)	388	(33)
2000	1474	677	(46)	162	(11)	94	(6)	541	(37)
2001	1294	608	(47)	151	(12)	72	(6)	463	(36)
2002	1130	586	(52)	133	(12)	78	(7)	333	(29)
2003	1314	529	(40)	152	(12)	80	(6)	553	(42)
2004	1656	726	(44)	132	(8)	110	(7)	688	(42)
2005	1187	713	(60)	119	(10)	143	(12)	212	(18)

Percentages are shown in brackets.

Appendix 4. Publications of the specialty: Health economics

Year	Total publications	Journal articles		Books		Working papers		Other publication types	
1991	625	413	(44)	161	(4)	14	(1)	743	(56)
1992	694	385	(58)	148	(5)	19	(1)	551	(50)
1993	709	393	(49)	173	(6)	21	(3)	645	(52)
1994	900	399	(58)	204	(4)	29	(2)	672	(52)
1995	889	438	(60)	198	(4)	31	(4)	652	(49)
1996	995	472	(68)	182	(3)	48	(3)	777	(53)
1997	1121	536	(65)	194	(3)	49	(4)	631	(45)
1998	1224	607	(67)	189	(2)	49	(4)	554	(40)
1999	1346	619	(72)	140	(3)	39	(5)	388	(33)
2000	1564	677	(59)	162	(3)	94	(7)	541	(37)
2001	1573	608	(66)	151	(2)	72	(5)	463	(36)
2002	1255	586	(63)	133	(3)	78	(7)	333	(29)
2003	1281	529	(58)	152	(5)	80	(8)	553	(42)
2004	1718	726	(66)	132	(3)	110	(8)	688	(42)
2005	1670	713	(71)	119	(3)	143	(11)	212	(18)

Percentages are shown in brackets.

Appendix 5. The 20 highest JIF ranking economics journals in JCR 2005

Journal name	JIF
Quarterly Journal of Economics	4.775
Journal of Economic Literature	4.054
Journal of Economic Geography	3.222
Journal of Health Economics	2.708
Journal of Economic Perspectives	2.634
Econometrica	2.626
Journal of Economic Growth	2.577
Journal of Financial Economics	2.385
Journal of Political Economy	2.245
Brookings Papers on Economic Activity	2.118
Journal of Risk and Uncertainty	2.100
Review of Economic Studies	2.035
Health Economics	1.919
Journal of Accounting & Economcis	1.877
American Economic Review	1.806
Economic Geography	1.757
Journal of International Economics	1.667
Journal of Monetary Economics	1.661
Journal of Law & Economics	1.609
Feminist Economics	1.595

Appendix 6. Publications of the research tradition: Cognitive therapy

Year	Total publications	Journal articles		Books		Other publication types	
1991	179	91	(51)	64	(36)	24	(13)
1992	193	112	(58)	58	(30)	23	(12)
1993	221	124	(56)	69	(31)	28	(13)
1994	223	147	(66)	58	(26)	18	(8)
1995	225	131	(58)	83	(37)	11	(5)
1996	238	138	(58)	90	(38)	10	(4)
1997	223	149	(67)	69	(31)	5	(2)
1998	298	188	(63)	107	(36)	3	(1)
1999	103	88	(85)	14	(14)	1	(1)
2000	302	217	(72)	79	(26)	6	(2)
2001	359	291	(81)	68	(19)	0	(0)
2002	382	283	(74)	96	(25)	3	(1)
2003	446	397	(89)	45	(10)	4	(1)
2004	553	431	(78)	111	(20)	11	(2)
2005	493	424	(86)	59	(12)	10	(2)

Percentages are shown in brackets.

Appendix 7. Publications of the research tradition: Behavior Therapy & behavior modification

Year	Total publications	Journal articles		Books		Other publication types	
1991	348	247	(71)	52	(15)	49	(14)
1992	263	195	(74)	21	(8)	47	(18)
1993	381	225	(59)	91	(24)	65	(17)
1994	268	201	(75)	40	(15)	27	(10)
1995	246	197	(80)	42	(17)	7	(3)
1996	279	232	(83)	42	(15)	5	(2)
1997	273	227	(83)	44	(16)	2	(1)
1998	254	173	(68)	79	(31)	2	(1)
1999	95	78	(82)	16	(17)	1	(1)
2000	245	203	(83)	32	(13)	10	(4)
2001	269	221	(82)	48	(18)	0	(0)
2002	297	252	(85)	45	(15)	0	(0)
2003	289	275	(95)	12	(4)	2	(1)
2004	227	204	(90)	20	(9)	3	(1)
2005	239	194	(81)	41	(17)	4	(2)

Percentages are shown in brackets.

Appendix 8. Publications of the research tradition: Psychoanalytic therapy

Year	Total publications	Journal articles		Books		Other publication types	
1991	790	585	(74)	182	(23)	23	(3)
1992	793	634	(80)	143	(18)	16	(2)
1993	959	700	(73)	240	(25)	19	(2)
1994	930	772	(83)	140	(15)	18	(2)
1995	878	667	(76)	202	(23)	9	(1)
1996	909	682	(75)	209	(23)	18	(2)
1997	786	590	(75)	189	(24)	7	(1)
1998	766	643	(84)	115	(15)	8	(1)
1999	357	282	(79)	71	(20)	4	(1)
2000	778	677	(87)	93	(12)	8	(1)
2001	775	635	(82)	139	(18)	1	(0)
2002	816	685	(84)	131	(16)	0	(0)
2003	922	802	(87)	120	(13)	0	(0)
2004	945	794	(84)	142	(15)	9	(1)
2005	812	739	(91)	57	(7)	16	(2)

Percentages are shown in brackets.

Appendix 9. The 20 highest JIF ranking psychology journals in JCR 2005

Journal name	JIF
Behavioral and Brain Sciences	9.885
Annual Review of Psychology	9.784
Psychological Bulletin	9.746
Trends in Cognitive Sciences	9.155
Psychological Review	7.986
Advances in Experimental Social Psychology	7.000
American Psychologist	6.460
Monographs of the Society for Research in Child Development	5.667
Journal of Experimental Psychology – General	5.242
Journal of Clinical Psychiatry	5.038
Psychotherapy and Psychosomatics	4.966
Journal of Cognitive Neuroscience	4.533
Psychological Science	4.502
Journal of Abnormal Psychology	4.383
Journal of Personality and Social Psychology	4.211
Neuropsychologia	4.119
Journal of the American Academy of Child and Adolescent Psychiatry	4.113
Neurobiology of Learning and Memory	4.091
Journal of Consulting and Clinical Psychology	4.023
Cognitive Psychology	3.932

* Early results of this work were presented at the 3rd International COLLNET Conference, New Delhi, India, March 2007. This work was partly financed by a generous grant from the Nordic Research School in Library and Information Science (NORSLIS).
† Corresponding Author: Tove Faber Frandsen. Royal School of Library and Information Science. Birketinget 6, DK-2300 Copenhagen S. Phone: + 45 32 58 60 66, e-mail: tff@db.dk.